aps
ASSOCIATION FOR
PSYCHOLOGICAL SCIENCE

Data: Sharing Is Caring

Margaret C. Levenstein (iD) and Jared A. Lyle (iD)
Inter-university Consortium for Political and Social Research, University of Michigan

Abstract
Data sharing promotes scientific progress by permitting replication of prior scientific analyses and by increasing the return on the human and financial investments made in data collection. The costs of data sharing can be reduced through the implementation of best practices in data management across the research life cycle; this article provides specific guidance on these practices. The benefits of data sharing will be reaped when researchers who share their data are rewarded with citations and recognition of the intellectual value inherent in producing new scientific data.

Keywords
data sharing, data stewardship, reproducibility, replication, data preservation, data citation

Data sharing is garnering increasing support across the social and natural sciences, including within the field of psychology. As the Data Sharing Working Group of the American Psychological Association (2015) recently noted, sharing data not only has the straightforward, key benefits of increased openness, accountability, and scientific progress, but also

> allows geographically dispersed individuals and those with limited resources to investigate scientific questions of interest, enables replication of analyses for verifying empirical findings, and opens extant data to analysis with new, more powerful, or integrative techniques than were available at the time of collection. (paragraph 2)

Several other benefits have been extolled for decades. Data sharing encourages multiple perspectives, saves money through reuse, promotes new research, leads to many more publications based on the data than would result if they were not shared, and protects against "fabrication, falsification, or distortion" (National Research Council, 1985, p. 14; see also Pienta, Alter, & Lyle, 2010; Piwowar, Day, & Fridsma, 2007). Many governmental agencies (Holdren, 2013; National Institutes of Health, Office of Extramural Research, 2007; National Science Foundation, n.d.), charitable foundations (Bill & Melinda Gates Foundation, 2017; Wellcome, n.d.-b), professional associations (Lupia & Elman, 2014), and

scientific journals (Data Access & Research Transparency, 2016; *Nature*, 2016; Silva, 2014) now formally recognize these benefits by encouraging or even mandating data sharing.

Data sharing is a boon not only to the scientific community, but also to the original investigator. Sharing encourages better data management, which improves the efficiency and quality of output (Academia Stack Exchange, n.d.). Sharing can provide credit and recognition, and extend a researcher's reach, much as publication does. When a researcher shares via an established and trusted archive, sharing reduces busy-work and information loss after a project is completed by transferring dissemination and user-support responsibilities to the repository.

Resistance to Sharing Data

Although the benefits of data sharing are numerous, the practice of data sharing is still not standard in many research communities, including the psychology community. One study, for instance, found that 73% of psychologists whose work was published in four top American Psychological Association journals did not

Corresponding Author:
Margaret C. Levenstein, University of Michigan, ICPSR, Ann Arbor, MI 48106-1248
E-mail: maggiel@umich.edu

share their data upon request, even though "all corresponding authors had signed a statement that they would share their data for such verification purposes" (Wicherts, Bakker, & Molenaar, 2011, Methods, paragraph 1; see also Wicherts, Borsboom, Kats, & Molenaar, 2006). Why the hesitation? Some researchers fear loss of control or being scooped. In practice, however, the original data producer has a first-mover advantage for publication, given his or her deep knowledge of the data. Sharing data increases the likelihood that they will make a significant impact through secondary analysis and makes possible future citations that increase the value of the data and publications related to them (LeClere, 2010).

Other researchers worry that their work will not be replicated—never mind that reproducibility is a core component of science. As the Open Science Collaboration (2015) has argued, "scientific claims should not gain credence because of the status or authority of their originator but by the replicability of their supporting evidence" (p. 943). It is not surprising, at an emotional or individual level, that researchers are loathe to take actions that would allow someone to find that they made a mistake, or that their claim does not hold up. However, this understandable fear undermines scientific progress. If researchers think their research is worth doing—that the knowledge is worth knowing, not just to advance careers or build curriculum vitae, but because humankind is better off with better knowledge, a better understanding of why the sun shines or people get sick or unhappy or find jobs—then they should share their data to ensure that the knowledge they produce is valid. CAVATICA, a data-analysis and -sharing platform for pediatric cancer research, illustrates this point; see http://www.cavatica.org/): When research is driven by cancer patients and their families (in this case, represented by CAVATICA's funder, the Dragon Master Foundation), concerns about data sharing are set aside in the interests of more rapid scientific discovery.[1] Rather than reinforce or reward tentative behavior, the academic and scientific community should acknowledge the social value of scientific progress by rewarding those researchers who accept the risk of being proven wrong by sharing their data.

Still other researchers worry about freeloading: "Why should I go to the effort to obtain grants and collect my own data if I am then required to share my data with others?" (Firebaugh, 2007, p. 207). Researchers need to be willing to make costly investments in collecting primary data, but there are ways to reward them that are consistent with scientific advance. The U.S. patent system is based on the principle that the original inventor gets credit in return for transparency. The scientific community should approach researchers'

concerns regarding individual "property rights" over data in the same way. We acknowledge that researchers want to receive credit as well as to advance knowledge, and that research subjects want their confidentiality to be protected as well as to advance knowledge. The interests of both researchers and subjects are completely consistent with sharing data, as long as confidentiality protections and requirements for data citation are implemented. It is those citations that provide researchers with the individual incentive, beyond the intrinsic value of advancing science, to produce and share data.

Incentivized Data Sharing

Important changes in institutional, disciplinary, journal, and funder practices are necessary to support individual researchers in sharing their data and incentivize them to do so. Academic organizations—departments, colleges and universities, disciplinary associations such as the Association for Psychological Science—should treat data sharing (or published documentation or description of data) as a productive, scientific contribution in and of itself, equivalent to a refereed academic publication. The sharing of data should be considered explicitly in tenure and promotion decisions. Doing so would both provide increased incentive for sharing and contribute to the social validation of sharing.

Journals should require, rather than simply recommend, data archiving. One recent study found that the odds of finding the data online increase almost 1,000-fold when journals mandate data archiving, compared with when they have no archiving policy (Vines et al., 2013). The same study found archiving rates to be "only very slightly higher" (p. 1304) at journals that recommend (but do not require) archiving than at journals with no policy. Without mandates, data simply are not shared. Journals should also require data citation, so that those researchers who do share their data receive credit. Data sets should be cited just as publications are, so that the data producers receive comparable credit for their contribution.

Similarly, funders should require data sharing by awardees who are producing valuable new data resources, and they should provide funding for sustainable access to data that are shared. Although the requirement to submit data-management plans has increased awareness of the value of data in the research community, the returns to funders' investments in the creation of new data can be fully realized only by taking the next step by requiring, enforcing, and funding sustainable plans for sharing data, and funders should withhold funding until sharing is verified. The National Institutes of Health in the United States and the Wellcome Trust in the United Kingdom already withhold a

portion of funding when researchers do not make their publications publicly available (van Noorden, 2014). The Wellcome Trust, for instance, withholds the final 10% of the total grant funding if publication-sharing requirements are not met (Wellcome, n.d.-a). The result of this policy change was a marked increase in compliance with Wellcome's open-access policy: The compliance rate rose to 91% in 2016, a significant improvement from 74% in 2015 (Wellcome, n.d.-c).

Increasing Researchers' Capacity for Data Sharing

Two common data-sharing concerns we hear from our own community of users at the Inter-university Consortium for Social and Political Research (ICPSR) at the University of Michigan involve the level of effort required of researchers and disclosure risk. The first concern is that data sharing is a lot of work. The second concern is that data cannot be shared because they are confidential or the study's subjects are identifiable. The latter is a particular concern when the data come from detailed, longitudinal surveys or administrative or other naturally occurring data. Solutions to each of these concerns are available and straightforward to implement.

As regards the first concern, it is true that data sharing does take effort. As one scientist-blogger (and, it should be noted, open-science enthusiast) wrote in exasperation, preparing data for sharing can be "a major pain in the ass and is really expensive, in terms of both the money and amount of time required" (Bruna, 2014, paragraph 5). Yet, with sufficient preparation, the effort needed to share data can be built into the standard data workflow, so that the eventual data sharing is nearly painless. As one data-sharing advocate noted, "Collecting data is a bit like cooking a good meal. If you clean as you go, when you are full and sleepy you will have much less to do" (LeClere, 2010, paragraph 12).

Guidance about good data-management practice is plentiful. Several data repositories have provided step-by-step instructions, including ICPSR's (2012b) "Guide to Social Science Data Preparation and Archiving" and DataONE's "Data Management Guide for Public Participation in Scientific Research" (Wiggins et al., 2013). Recent publications have condensed the guidance into easily digestible rules of thumb, such as "Ten Simple Rules for the Care and Feeding of Scientific Data" (Goodman et al., 2014). This guidance typically addresses four primary components of data management:

- Planning: Many federal agencies now require that formal data-management plans be submitted with research proposals. A good plan "describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyze, and store those data, and what mechanisms you will use at the end of your project to share and preserve your data" (Stanford Libraries, n.d., paragraph 1). Planning streamlines your workflow, prepares you for what to expect, and helps you avoid problems. ICPSR (2012a) provides step-by-step guidelines for creating a data-management plan (see Table 1 for a summary).

- Standardization: A well-prepared data collection includes all information that secondary users will need to make sense of the data. The simplest way to meet this goal is to standardize documentation as it is created. Make sure to document everything! This includes labels, values, code, and the process used to generate the final data product. Workflow managers such as the Open Science Framework (https://osf.io/), GitHub (https://github.com/), and SEAD (http://sead-data.net/) can help you manage active projects. Table 2 presents useful questions to consider when planning how to standardize data.

- Description: Create a codebook or user guide explaining the overall data collection, including the purpose, format, and methodological details (ICPSR, 2011). Good descriptions help answer questions about "the why, who, what, when, where, and how of the project" (Wiggins et al., 2013, p. 7). This information will help secondary users understand how to use the data, including how to replicate the study. That same information can be useful when creating an online finding aid for other researchers who are studying areas covered by the data and searching for relevant data. Table 3 summarizes the different elements you should make sure to include in your codebook.

- Archiving: Some repositories will help review and enhance a data collection prior to publication. A trusted archive should provide a landing page with a persistent, globally unique identifier so that other people can find and cite your data, and provide the long-term digital support needed to preserve and disseminate the data well beyond the end of your formal research project. For example, Figure 1 shows a standard study home page at ICPSR. In addition to the study's title, the names of individuals responsible for the production of the data, and links to the data and other information, it provides a standard data citation, so that any secondary users know how the data should be cited. There are now available many data repositories that will preserve and share data from individual researchers. These include general data repositories, such as Dryad (http://

Table 1. Elements of a Data-Management Plan

Element	Description
Data description	Description of the information to be gathered, including its nature and scale
Existing data	Survey of existing data relevant to the project and discussion of whether and how these data will be integrated into the project
Format	The formats in which the data will be generated, maintained, and made available, including a justification for the procedural and archival appropriateness of those formats
Metadata	Description of the metadata to be provided along with the generated data and discussion of the metadata standards used
Storage and backup	Storage methods and backup procedures for the data, including the physical and cyber resources and facilities that will be used
Security	Description of the technical methods and procedures that will be used to protect information, including confidential information, and how permissions, restrictions, and embargoes will be enforced
Responsibility	Names of the individuals responsible for data management
Intellectual property rights	Names of the entities or persons who will hold the intellectual property rights to the data; description of how intellectual property will be protected, if necessary, and of any copyright constraints (e.g., copyrighted data-collection instruments)
Access and sharing	Description of how data will be shared, including access procedures, embargo periods, technical mechanisms for dissemination, and whether access will be open or granted only to specific user groups; the time frame for data sharing and publishing
Audience	Description of potential secondary users of the data
Selection and retention periods	Description of how data will be selected for archiving, how long the data will be held prior to archiving, and plans for eventual destruction or termination of the data collection in the future
Archiving and preservation	Procedures in place or envisioned for long-term archiving and preservation of the data, including succession plans for the data should the expected archiving entity go out of existence
Ethics and privacy	Discussion of how informed consent will be handled and how privacy will be protected, including any exceptional arrangements that might be needed to protect subjects' confidentiality and other ethical issues that may arise
Budget	Costs of preparing data and documentation for archiving and how these costs will be paid; these costs may be included in requests for funding
Data organization	Description of how data will be managed during the project, including information about version control and naming conventions
Quality assurance	Procedures for ensuring data quality during the project
Legal requirements	Listing of all relevant federal or funder requirements for data management and data sharing

Note: The recommendations in this table are drawn from Inter-university Consortium for Political and Social Research (2012a).

datadryad.org/), figshare (https://figshare.com), and Zenodo (https://zenodo.org), as well as disciplinary archives, such as ICPSR, and local institutional repositories typically run by a university library (Vitale et al., 2017).

Effective data management takes practice, but the payoff is rewarding. That same scientist-blogger who complained about data sharing being "a major pain in the ass" later wrote, "It's also important for me to remember that it will get cheaper and cheaper every time I do it (e.g., preparing metadata was a snap this time because I used the template from my prior submissions . . .)" (Bruna, 2014, paragraph 5).

The second concern we hear from ICPSR users is more complicated, but still manageable. There is justifiable reason to guard against disclosure of confidential or potentially identifiable data, particularly when the data are detailed or longitudinal. Even if basic identifying information is suppressed, the large amount of information now available at low cost in public and commercial data sets makes re-identification easier than in the past (Bradley & Sweeney, 2004). Researchers often want access to data (e.g., detailed geographic information) that makes other data easily re-identifiable, or to personally identifiable information itself, for linkage to, for example, electronic health records or administrative records on health-care expenditures (e.g., records from

Table 2. Questions to Consider When Planning to Standardize Data

Metadata topic	Questions
Data and file structure	What is the data file going to look like and how will it be organized? What is the unit of analysis?
Naming conventions	How will files and variables be named? What naming conventions will be used to achieve consistency?
Data integrity	How will data be input or captured? Will numeric or character formats be used? What checks will be used to find invalid values, inconsistent responses, incomplete records, etc.? What checks will be used to manage the data versions as the project proceeds through data entry, cleaning, and analysis?
Construction of variables	What variables will be constructed following the collection of the original data? How will these be named and documented?
Project documentation	What steps will be taken to document decisions that are made as the project unfolds? How will information on field procedures, coding decisions, and construction of variables be recorded?

Note: This table was adapted with permission from *Research Connections* (2017, p. 5).

the Centers for Medicare & Medicaid Services). Other data, such as videos of children, simply lose too much value for many studies if noise is added to obscure recognition (Gilmore & Adolph, 2017). Researchers often assume that the interest of the individuals participating in a study is best protected by not sharing. But participants share their time and personal information because they want scientific progress. Researchers owe it to study participants, and to the taxpayers and foundations that support research, to extract as much scientific knowledge as possible from the support that they have given to these research endeavors.

There are now multiple straightforward, accessible ways to address concerns about sharing data that contain sensitive and confidential information. In some cases these issues are best addressed with the assistance of a qualified data archive. Here, we describe several of the steps that you can take to address these concerns at different stages of a research project:

- Use appropriate consent protocols: At the beginning of a study, it is important to make use of consent protocols that explicitly permit sharing, with confidentiality protection, for research purposes. Although it is essential that confidentiality is maintained, proper informed-consent language should not unduly restrict eventual data sharing. For example, avoid language that promises to limit data access to a particular research team or indicates that the consent will expire after a period of time. ICPSR (2017c) maintains an online resource of recommended informed-consent language for data sharing (see Table 4).

- Separate personally identifiable information: If a data set includes personally identifiable information, it is recommended that these variables be separated from the rest of the data set and multiple crosswalks between the files be created and encrypted in order to protect private information without having to destroy it. First, give each

Table 3. Guidelines for Describing Data

Documentation topic	Issues to consider
Design documentation	Consider what the documentation or metadata will look like, how they will be produced, and what documentation standard (e.g., DDI) will be used. See www.ddialliance.org for more information on the DDI standard.
Variable-level description	Description at this level should include the names of all variables and how they are labeled in the files, the text of all questions, values of the variables and how they are coded in the files, summary statistics, details on missing data, skip patterns, information on how the variables were constructed, and notes (Inter-university Consortium for Political and Social Research, 2011).
Collection-level description	Description at this level should include the name and affiliation of each principal investigator, the title of the collection, the names of any funding sources, the names of all individuals who collected or produced the data, a description of the project, a description of the sample and sampling procedures and weights, a description of all sources of the data, the units of analysis, a list of any related publications, technical information on the files, and information on the data-collection instruments used (*Research Connections*, 2017).

Fig. 1. Example of the study home page for a data collection archived at the Inter-university Consortium for Political and Social Research.

subject a study-specific unique ID. Second, create a crosswalk between the personally identifiable information and a new, otherwise completely meaningless, identification key. Third, create a second crosswalk between that "meaningless" identification key and your study's IDs. Fourth, create a file that contains the personally identifiable information and the arbitrary IDs. Encrypt it. Finally, create a second file that contains the arbitrary IDs and the study IDs. Encrypt it. Keep your study data with the study IDs; keep the personally identifiable information separate. The file with your study data will be the basis for the

research data sets that are shared with the general research community. The encrypted files should be kept in a secure computing environment where they may be accessed by qualified researchers who want to enhance and analyze the data with other information (e.g., data from subsequent waves of a survey or administrative data, such as those in the National Death Index).

• Anonymize and desensitize the data: Public-use versions of the original data can be created by removing direct personal identifiers (e.g., names, identification numbers) as well as other variables that increase the likelihood of re-identification

Table 4. Model Informed-Consent Language for Data Sharing

Sample 1.	Study staff will protect your personal information closely so no one will be able to connect your responses and any other information that identifies you. Federal or state laws may require us to show information to university or government officials (or sponsors), who are responsible for monitoring the safety of this study. Directly identifying information (e.g. names, addresses) will be safeguarded and maintained under controlled conditions. You will not be identified in any publication from this study.
Sample 2.	The information in this study will be used only for research purposes and in ways that will not reveal who you are. Federal or state laws may require us to show information to university or government officials (or sponsors) who are responsible for monitoring the safety of this study. You will not be identified in any publication from this study.

Note: These examples are taken from the Model Language section of Inter-university Consortium for Political and Social Research (2017c).

(e.g., home address or detailed geography, exact measures of income). This option provides secondary users with access to data, albeit data that are missing potential variables of interest. Rather than removing variables, it may be possible to recode them. Examples of recoding include converting exact birth dates to time intervals or exact income levels to income ranges (ICPSR, 2017a). Alternatively, you may use statistical methods of protecting confidentiality that "blur or disguise the data in such a way that individual data items cannot be uniquely associated with or attributed to a particular person or establishment" (American Statistical Association, 2008, p. 4; see also Federal Committee on Statistical Methodology, 2005).

- Put the data behind a physical or virtual firewall: There are a number of alternative methods for providing researchers with access to confidential data for analytic purposes while still protecting it from disclosure. One option that works well for popular, well-documented data is software that permits on-line analysis and automatically checks the output for disclosure risk. Such software allows secondary users immediate analytic access to restricted-use data behind an interface that automates disclosure protections. For example, results with cell counts beyond a sensitivity threshold may be suppressed to protect against potential identification of subjects. Researchers cannot directly observe the underlying microdata. A second option is to make data available in a virtual data enclave. A virtual data enclave provides access to restricted data on a remote server, similar to a remote-desktop environment. Users cannot download or upload data, and all software needed for analysis is preinstalled in the secure environment. Output can be removed from the virtual enclave only by a third party and may be reviewed for disclosure risk before being moved. A researcher can see but cannot remove the microdata and also cannot combine the microdata with other data that might permit re-identification. Finally, data can be shared through a physical data enclave. In this case, approved users visit a physical location that is monitored and secured. They cannot bring electronic equipment into the enclave, and all output, notes, and other materials must be reviewed for disclosure risk before leaving the enclave (ICPSR, 2017b).

Data-sharing concerns can be overcome with sufficient planning. Tools and processes to assist researchers are in place, especially through qualified data archives.

Data sharing need not be overly burdensome, and confidential or identifiable data can be shared using appropriate access and control mechanisms. The benefits of data sharing far outweigh the costs. The research community can increase data sharing, and the benefits reaped, by implementing small changes in standard academic practice, including requiring citation of data sets and valuing data sharing as a scientific contribution comparable to publication. We owe it to research subjects and to the taxpayers and foundations that support research to extract as much scientific knowledge as possible from their support. Sharing data with other researchers is one way to achieve that.

Action Editor

Daniel J. Simons served as action editor for this article.

Author Contributions

M. C. Levenstein and J. A. Lyle jointly generated the ideas and text in this article. Their names are listed in alphabetical order.

ORCID iD

Margaret C. Levenstein https://orcid.org/0000-0002-9641-2725
Jared A. Lyle https://orcid.org/0000-0001-8623-7612

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Note

1. The title for this article is borrowed from the Dragon Master Foundation's swag. For more information about the Dragon Master Foundation's support for data sharing, see https://www.dragonmasterfoundation.org/.

References

Academia Stack Exchange. (n.d.). *How do open science practices affect tenure and promotion decisions?* Retrieved from https://academia.stackexchange.com/questions/51974/how-do-open-science-practices-affect-tenure-and-promotion-decisions

American Psychological Association, Data Sharing Working Group. (2015). *Data sharing: Principles and considerations for policy development.* Retrieved from http://www.apa.org/science/leadership/bsa/data-sharing-report.aspx

American Statistical Association. (2008). *Data access and personal privacy: Appropriate methods of disclosure control.* Retrieved from http://www.amstat.org/asa/files/pdfs/POL-DataAccess-PersonalPrivacy.pdf

Bill & Melinda Gates Foundation. (2017). *Bill & Melinda Gates Foundation open access policy.* Retrieved from https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy

Bradley, M., & Sweeney, L. (2004). How (not) to protect genomic data privacy in a distributed network: Using trail re-identification to evaluate and design anonymity protection systems. *Journal of Biomedical Informatics, 37*, 179–192.

Bruna, E. M. (2014, September 4). The opportunity cost of my #OpenScience was 36 hours + $690 (updated) [Web log post]. Retrieved from http://brunalab.org/blog/2014/09/04/the-opportunity-cost-of-my-openscience-was-35-hours-690/

Data Access & Research Transparency. (2016). *The Journal Editors' Transparency Statement (JETS)*. Retrieved from https://www.dartstatement.org/2014-journal-editors-statement-jets

Federal Committee on Statistical Methodology. (2005). *Report on statistical disclosure limitation methodology* (Statistical Policy Working Paper 22, second version). Retrieved from https://www.hhs.gov/sites/default/files/spwp22.pdf

Firebaugh, G. (2007). Replication data sets and favored-hypothesis bias. *Sociological Methods & Research, 36*, 200–209.

Gilmore, R. O., & Adolph, K. E. (2017). Video can make behavioural science more reproducible. *Nature Human Behavior, 1*(7), Article 0128. doi:10.1038/s41562-017-0128

Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., . . . Slavkovic, A. (2014). Ten simple rules for the care and feeding of scientific data. *PLOS Computational Biology, 10*(4), Article e1003542. doi:10.1371/journal.pcbi.1003542

Holdren, J. P. (2013). *Increasing access to the results of federally funded scientific research*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Inter-university Consortium for Political and Social Research. (2011). *Guide to codebooks*. Retrieved from http://www.icpsr.umich.edu/files/deposit/Guide-to-Codebooks_v1.pdf

Inter-university Consortium for Political and Social Research. (2012a). *Guidelines for effective data management plans*. Retrieved from https://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf

Inter-university Consortium for Political and Social Research. (2012b). *Guide to social science data preparation and archiving* (5th ed.). Retrieved from https://www.icpsr.umich.edu/files/deposit/dataprep.pdf

Inter-university Consortium for Political and Social Research. (2017a). *Confidentiality*. Retrieved from https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confidentiality/

Inter-university Consortium for Political and Social Research. (2017b). *Data enclaves*. Retrieved from https://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/enclave.html

Inter-university Consortium for Political and Social Research. (2017c). *Recommended informed consent language for data sharing*. (2017). Retrieved from https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confidentiality/conf-language.html

LeClere, F. (2010, August 3). Too many researchers are reluctant to share their data. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/Too-Many-Researchers-Are/123749

Lupia, A., & Elman, C. (2014). Openness in political science: Data access and research transparency: Introduction. *Political Science & Politics, 47*, 19–42. doi:10.1017/S1049096513001716

National Institutes of Health, Office of Extramural Research. (2007). *NIH data sharing policy*. Retrieved from https://grants.nih.gov/grants/policy/data_sharing/

National Research Council. (1985). *Sharing research data*. Washington, DC: National Academies Press.

National Science Foundation. (n.d.). *Dissemination and sharing of research results*. Retrieved from https://www.nsf.gov/bfa/dias/policy/dmp.jsp

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716. doi:10.1126/science.aac4716

Nature. (2016). *Announcement: Where are the data?* Retrieved from http://www.nature.com/news/announcement-where-are-the-data-1.20541

Pienta, A. M., Alter, G., & Lyle, J. A. (2010). *The enduring value of social science research: The use and reuse of primary research data*. Retrieved from http://hdl.handle.net/2027.42/78307

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLOS ONE, 2*(3), Article e308. doi:10.1371/journal.pone.0000308

Research Connections. (2017). *Guide to archiving data with* Research Connections: *Considerations throughout the research lifecycle*. Retrieved from http://www.researchconnections.org/files/childcare/pdf/guide-for-archiving-rc.pdf

Silva, L. (2014, February 26). PLOS' new data policy: Public access to data [Web blog post]. Retrieved from http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2/

Stanford Libraries. (n.d.). *Data management plans*. Retrieved from https://library.stanford.edu/research/data-management-services/data-management-plans

van Noorden, R. (2014). Funders punish open-access dodgers. *Nature, 508*, 161. doi:10.1038/508161a

Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., . . . Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal, 27*, 1304–1308. doi:10.1096/fj.12-218164

Vitale, C. R., Carlson, J., Hodge, A. E., Hswe, P., Johns, E., Johnston, L. R., . . . Zilinski, L. (2017). Making the case for institutional data repositories. In L. R. Johnston (Ed.), *Curating research data: Vol. 2. A handbook of current practice* (pp. 158–161). Chicago, IL: Association of College and Research Libraries.

Wellcome. (n.d.-a). *Complying with our open access policy*. Retrieved from https://wellcome.ac.uk/funding/managing-grant/complying-our-open-access-policy

Wellcome. (n.d.-b). *Sharing research data to improve public health: Full joint statement by funders of health research*.

Retrieved from https://wellcome.ac.uk/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health

Wellcome. (n.d.-c). *Wellcome and COAF open access spend 2015-16*. Retrieved from https://wellcome.ac.uk/fund ing/managing-grant/wellcome-and-coaf-open-access-spend-2015-16

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results.

PLOS ONE, *6*(11), Article e26828. doi:10.1371/journal .pone.0026828

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*, 726–728.

Wiggins, A., Bonney, R., Graham, E., Henderson, S., Kelling, S., LeBuhn, G., . . . Weltzin, J. (2013). *Data management guide for public participation in scientific research*. Retrieved from https://www.dataone.org/sites/all/ documents/DataONE-PPSR-DataManagementGuide.pdf